Project Completion

Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

PI: John L Schnase, GSFC

Attendees at the 4th Invasive Species Forecasting System Science Team meeting held in May 2004 in Fort Collins, Colorado, at the USGS Fort Collins Science Center. Attendees represented: NASA (HQ, ARC, GSFC, JPL, SSC), USGS (HQ, FCSC, NBII), University of Alabama at Huntsville, Colorado State University, Mississippi State University, and San Francisco State University.

Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

PI: John L Schnase,GSFC

Objective

- Develop an infrastructure for modeling the changing geospatial distribution of the living components of the Earth environments.
- Convert existing biotic geospatial modeling and prediction environment to cluster computing systems with high scalability and performance

The Grand Staircase-Escalante National Monument in southern Utah is one of the study sites for applying geospatial models to predicting how invasive plant species spread.

Approach

- Analyze current modeling system for parallel implementation issues
- Parallelize the system, maximizing scalability and performance
- Design and implement efficient data ingestion mechanisms to maintain performance on large models

Co-I's

Tom Stohlgren,United States Geological Survey (USGS),
Robin Reich, Colorado State University (CSU),
Jeff Pedelty, Jim Smith,GSFC

Key Milestones

- Baseline "Plant Diversity" code on 2 test problems and analyze performance issues for parallelization 7/02
- Parallelize and demonstrate "Plant Diversity" code performance and scaling 7/03
- "Plant Diversity" code ingests 10x more data than baseline and demonstrates predicted performance improvement 7/04
- Customer delivery - Achieve sustained use of "Plant Diversity" code by USGS and CSU customers (renegotiated 6/04) 12/04

TRL_{in}=4

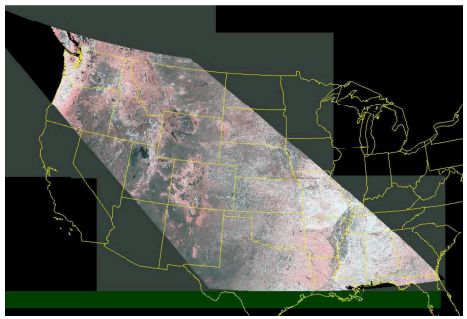
Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

PI: John L Schnase, GSFC

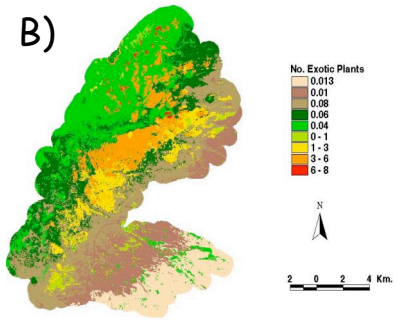
Objective

- Develop an infrastructure for modeling the changing geospatial distribution of the living components of the Earth environments.
- Convert existing biotic geospatial modeling and prediction environment to cluster computing systems with high scalability and performance

A)



B)



A) Continental scale MODIS time series data produced by the Invasive Species Forecasting System's data service B) Predictive map of exotic species richness in the Cerro Grande Fire Site produced by ISFS's analysis and modeling service. "A" is one of the most important predictor variables for the model that produces "B"

Accomplishments

- Exceeded all stated objectives !
- Completed parallel statistical algorithm improvements meeting specified performance improvements for target sites. The 'adaptive kriging' solution exceeds original performance goal. Documented process improvements and scaling curves.
- Deployed at USGS an operational, web-based modeling application based on an extensible and standard framework. The application provides the capability to specify run parameters, launch model runs and generate persistant predictive models and maps for designated study sites. Application includes parallel algorithm code base.
- Authored and delivered engineering documentation: Requirements, Design, Test Plan, User Guide, Configuration and Maintenance Manual. Conducted user and partner workshops to demonstrate new modeling capabilities and perform collaborative discovery. Documentation publicly available at http://tamarisk.sesda.com/summer2003/CT_deliverables.html
- Specified and delivered to agency partner USGS, Apple Xserve cluster server. Trained agency IT on system admin.
- Used algorithms and data to produce the first National Tamarisk Habitat Suitability Map for the US

Impact *(see following charts)*

- Project became the the basis for USGS/NASA Invasive Species Forecasting System national application.

Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

Technology Accomplishments

Scalable processing improvements with Cerro Grande Fire Site (CGFS) data

Re-engineered original S-plus code into a Fortran routine ⇨

- ***Reduced processing from 18 days to 61 mins***

Parallelized Fortran code ⇨

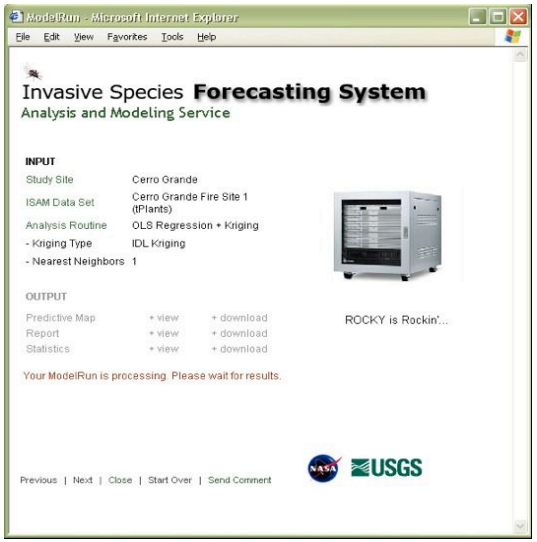
- ***Reduced processing from 61 mins to 2.47 mins***

18 days ⇨ 2.5 min changes the science!

Performance of "Adaptive Kriging" for the CGFS study site exceeded goals ⇨

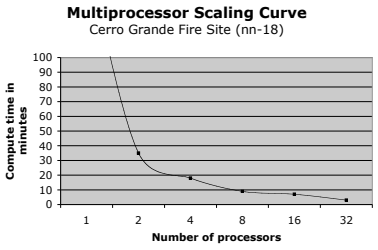
- *1x Area: goal of 2.47 min, achieved 33 seconds. Exceeded goal by 4.5x*
- *10x Area: goal is 24.7 min, achieved 4 min 2 sec. Exceeded goal by 6x*

Dramatically improved both the quality and capacity of science results for our USGS clients through *code optimization and cluster computing*



"Constraints in computational time often forced us to substitute simple models for complex, more realistic and accurate models. We needed to greatly reduce computational time to allow us to evaluate larger areas more quickly."

- Tom Stohlgren, Director National Institute of Invasive Species Science, USGS



Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

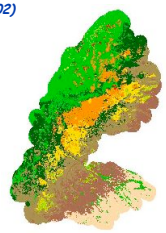
Science Accomplishments

New MODIS Time Series data products developed and deployed in this project improve the accuracy of species distribution predictions

MODIS NDVI* Time Series

Cerro Grande Fire Site

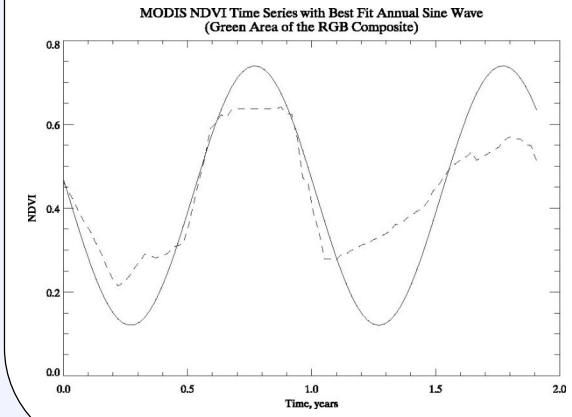
(Spring 2000 - Spring 2002)



Analysis by:

Jeff Morisette & Jeff Pedelty

NASA Goddard Space Flight Center

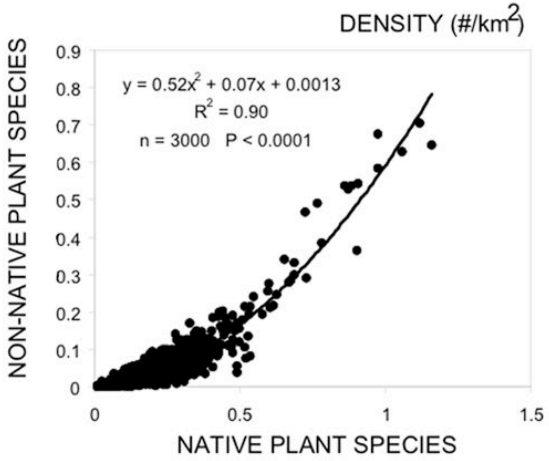


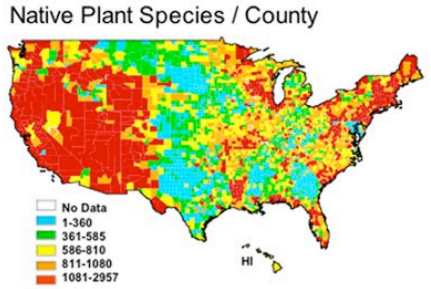
* Normalized Difference Vegetation Index, useful for assessing the health and density of vegetation

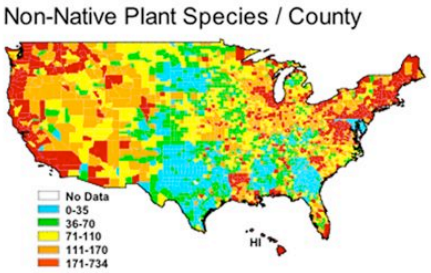
Improved data access and capacity for continental-scale modeling is enabling a new scientific understanding of global patterns of invasion …

A new "The Rich Get Richer" theory - developed by CO-I Tom Stohlgren (USGS) - shows that areas of high biodiversity are also highly vulnerable to invasions. Refutes long-standing notion that high species richness protects environments from invasion …

Invasive plants like high N, H_2O, sunlight just like native plants …



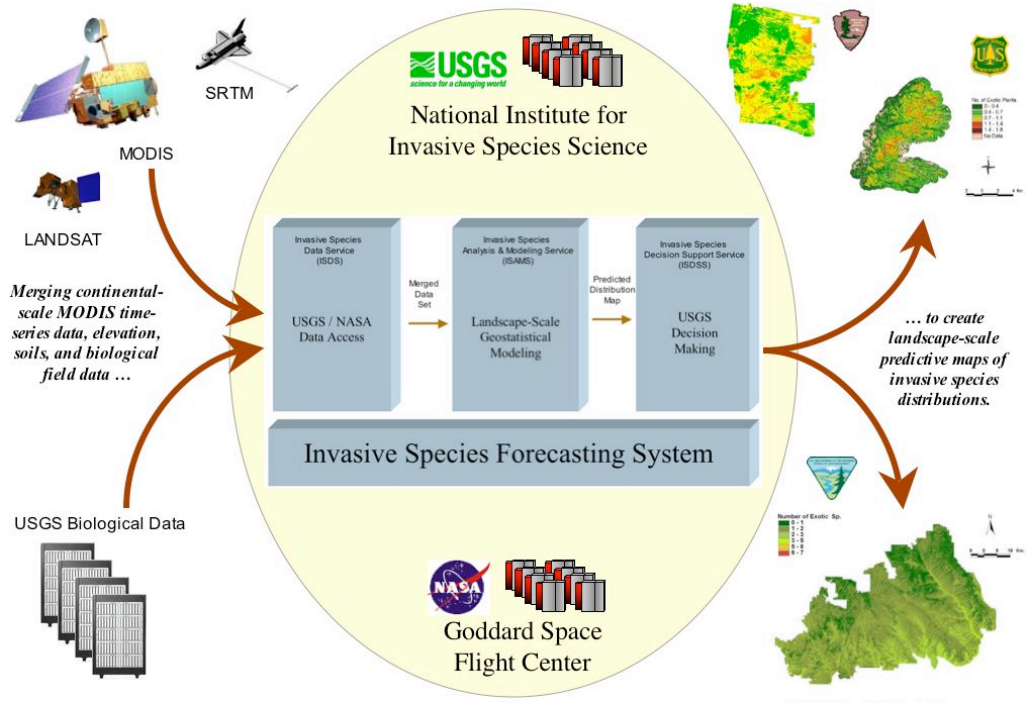


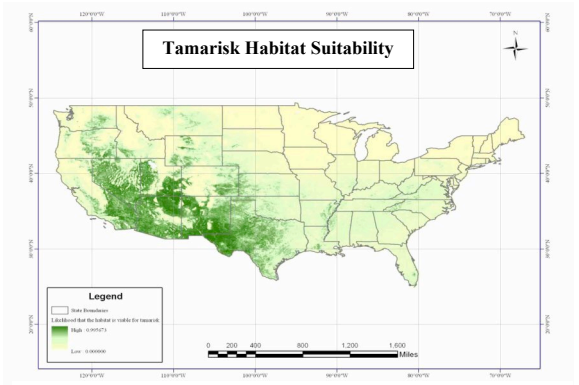


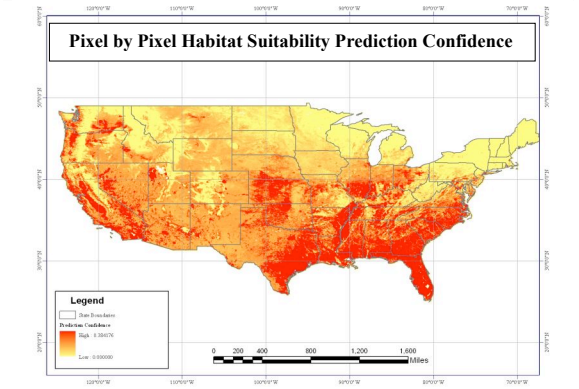
Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

Science Accomplishments → **Tech Transfer and Application Impact**

Data and algorithms used to produce the first "National Tamarisk Habitat Suitability Map" for the Continental United States







For the first time the land management community can see continental scale patterns of invasion!

The work started with the CT award has become the basis for the NASA/USGS Invasive Species Forecasting System (ISFS) national application !
http://InvasiveSpecies.gsfc.nasa.gov

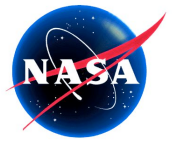
The ISFS is becoming a foundation technology in USGS's Early Detection Rapid Response mission !

Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

A $12.5M investment in Invasive Species Forecasting has grown out of the CT award:

Proposals Written and Won

2006 – 09	*Using the USGS Invasive Species Forecasting System to Support National Park Service Decisions on Fire Management Activities and Invasive Plant Species Control*, sponsored by NASA Science Mission Directorate, ($500,000), (Co-I, with J.T. Morisette (PI), N. Bensen, & B. Welch).
2004 – 07	*Fingerprinting Native and Non-Native Biodiversity in the Western United States*, sponsored by NASA Office of Earth Science, Science Division, ($1,200,000), (Co-I, with T.J. Stohlgren (PI) & J.T. Morisette).
2004 – 07	*Value Added Products from Vegetation and Precipitation Time-Series Data Sets in Support of Invasive Species Prediction*, sponsored by NASA Office of Earth Science, Science Division, ($646,700), (Co-I, with J.T. Morisette (PI), & T.J. Stohlgren).
2003 – 08	*The Invasive Species Data Services: Towards Operational Use of ESE Data in the USGS Invasive Species DSS*, sponsored by NASA Office of Earth Science, Applications Division, ($3,112,800), (PI, with J.A. Smith & T.J. Stohlgren).
2003 – 08	*ITR: Science on the Semantic Web – Prototypes in Biodiversity Informatics*, sponsored by the National Science Foundation Information Technology Research Program, ($3,750,000), (Co-I, with T.W. Finin (PI), J.A. Hendler, & J.F. Quinn).
2002 – 05	*Biotic Prediction: Building the High Performance Computing and Communications Infrastructure for Public Health and Environmental Forecasting*, sponsored by NASA Earth Science Technology Office, Computational Technologies Program, ($975,000), (PI, with J.A. Smith & T.J. Stohlgren).
2001 – 05	*Predicting Regional-Scale Exotic Plant Invasions in Grand Staircase-Escalante National Monument*, sponsored by NASA Earth Science Applications Directorate, Carbon Cycle Science Applications Program, ($715,000), (PI, with J.A. Smith & T.J. Stohlgren).
2001 – 04	*The GeoGrid Project*, sponsored by National Science Foundation, Information Technology Research Program, ($750,000), (Co-I, with P. Agouris).
2001 – 02	*Application of Internet Supercomputing Resources to USGS Research Questions*, sponsored by US Geological Survey, ($100,000), (Co-I, with A. Frondorf (PI), & T.J. Stohlgren).



Achievement Quads

Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

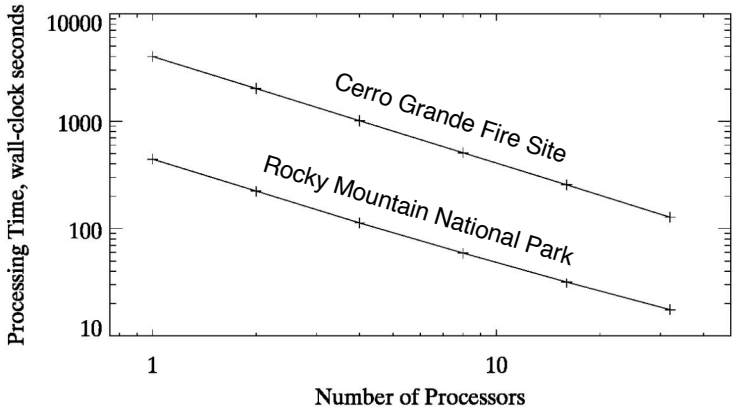
PI: John L Schnase, GSFC

Description

- Convert existing biotic geospatial modeling and prediction environment to cluster computing systems with high scalability and performance

Objective

- Infrastructure for modeling the changing geospatial distribution of the living components of the Earth environments.



Test case scaling curves on GSFC Medusa Linux cluster

Accomplishments

- Through parallelization, improved performance of the kriging routine, a spatial interpolator, which is the major computational bottleneck.
- Demonstrated result on two standard study sites that represent real operational scenarios:
  - Cerro Grande Fire Site in Los Alamos
  - Rocky Mountain National Park, Colorado
- Achieved 25X speed up on 32-processors.
- Updated Software Requirements document, Software Design document, and Test Plan.
- Provided Software Requirements Trace Matrix.

Key Milestones *(10 milestones total)*

E - Baseline *PlantDiversity* code on 2 test problems and analyze performance issues for parallelization 7/02

F - Parallelize and demonstrate *PlantDiversity* code performance and scaling 7/03

G - *Plant Diversity* code ingests 10x more data than baseline and demonstrates predicted performance improvement 7/04

K - Customer delivery - Achieve sustained use of *Plant Diversity* code by ten customers. 12/04

$TRL = 4_{in} - 4_{current}$

Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

PI: John L Schnase, GSFC

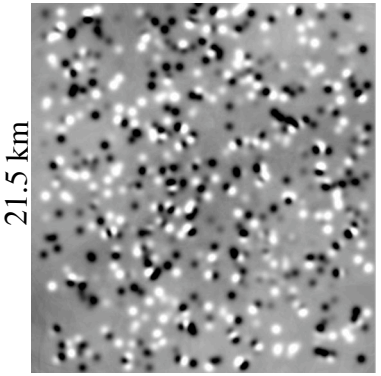
Description

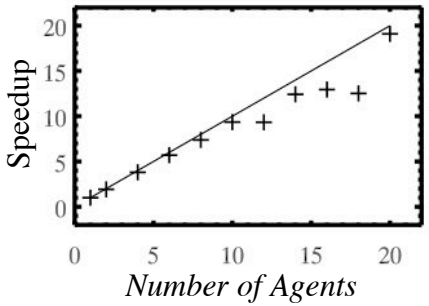
- Convert existing biotic geospatial modeling and prediction environment to cluster computing systems with high scalability and performance

Objective

- Infrastructure for modeling the changing geospatial distribution of the living components of the Earth environments.

Synthetic dataset for the Cerro Grande Wildfire Site
(700+ randomly spaced points)





Accomplishments

- Developed 'adaptive kriging' solution that exceeds milestone goal. Benchmarks run on new Apple G5 Xserve cluster, provided by Milestone-O1.
- Benchmark-1: 790 pts, 79 nn, 1x area
  - Size is 652 x 715, 30m pixels
  - Goal is 2.47 min, achieved 33 seconds (40 jobs)
  - Exceeded goal by 4.5x
- Benchmark-2: 790 pts, 79 nn, 10x area
  - Implemented as 1x physical area with 10x resolution, i.e. √10 bigger in each dim (2062 x 2261, 9.5m pixels)
  - Goal is 24.7 min, achieved 4 min 2 sec (40 jobs)
  - Exceeded goal by 6x
- Delivered updates to Requirements, Design, and Test Documents. Delivered initial User's Guide. Made documented source code publicly available via the Web.

Key Milestones *(10 milestones total)*

E - Baseline *PlantDiversity* code on 2 test problems and analyze performance issues for parallelization 7/02

F - Parallelize and demonstrate *PlantDiversity* code performance and scaling 7/03

G - *Plant Diversity* code ingests 10x more data than baseline and demonstrates predicted performance improvement 7/04

K - Customer delivery - Achieve sustained use of *Plant Diversity* code by USGS and CSU customers (reneg 6/04) 12/04

$TRL=4_{in}-5_{current}$

Schnase, Biotic Prediction: HPCC Infrastructure for Public Health and Environmental Forecasting

PI: John L Schnase, GSFC

Description

- Convert existing biotic geospatial modeling and prediction environment to cluster computing systems with high scalability and performance

Objective

- Infrastructure for modeling the changing geospatial distribution of the living components of the Earth environments.

End-users at USGS use the project's web application to evaluate models and generate predictive maps

Accomplishments

- Demonstrated predictive modeling application and utilities at USGS hosted science conference
- Delivered updates to Requirements, Design, and Test Documents. Delivered initial User's Guide. Made documented source code publicly available via the Web.
- Delivered Maintenance Manual for as-built system and trained on-site IT on system installation and maintenance
- Made documented source code publicly available via the Web.

Key Milestones *(10 milestones total)*

E - Baseline *PlantDiversity* code on 2 test problems and analyze performance issues for parallelization 7/02

F - Parallelize and demonstrate *PlantDiversity* code performance and scaling 7/03

G - *Plant Diversity* code ingests 10x more data than baseline and demonstrates predicted performance improvement 7/04

K - Customer delivery - Achieve sustained use of *Plant Diversity* code by USGS and CSU customers (reneg 6/04) 12/04

$TRL=4_{in}-6_{current}$